■ SPINE

Machine-learning models for the prediction of ideal surgical outcomes in patients with adult spinal deformity

D. Wang,
Q. Wang,
P. Cui,
S. Wang,
D. Han,
X. Chen,
S. Lu

From Xuanwu Hospital of Capital Medical University, National Clinical Research Center for Geriatric Diseases, Beijing, China

Aims
Adult spinal deformity (ASD) surgery can reduce pain and disability. However, the actual surgical efficacy of ASD in doing so is far from desirable, with frequent complications and limited improvement in quality of life. The accurate prediction of surgical outcome is crucial to the process of clinical decision-making. Consequently, the aim of this study was to develop and validate a model for predicting an ideal surgical outcome (ISO) two years after ASD surgery.

Methods
We conducted a retrospective analysis of 458 consecutive patients who had undergone spinal fusion surgery for ASD between January 2016 and June 2022. The outcome of interest was achievement of the ISO, defined as an improvement in patient-reported outcomes exceeding the minimal clinically important difference, with no postoperative complications. Three machine-learning (ML) algorithms – LASSO, RFE, and Boruta – were used to identify key variables from the collected data. The dataset was randomly split into training (60%) and test (40%) sets. Five different ML models were trained, including logistic regression, random forest, XGBoost, LightGBM, and multilayer perceptron. The primary model evaluation metric was area under the receiver operating characteristic curve (AUROC).

Results
The analysis included 208 patients (mean age 64.62 years (SD 8.21); 48 male (23.1%), 160 female (76.9%)). Overall, 42.8% of patients (89/208) achieved the ideal surgical outcome. Eight features were identified as key variables affecting prognosis: depression, osteoporosis, frailty, failure of pelvic compensation, relative functional cross-sectional area of the paraspinal muscles, postoperative sacral slope, pelvic tilt match, and sagittal age-adjusted score match. The best prediction model was LightGBM, achieving the following performance metrics: AUROC 0.888 (95% CI 0.810 to 0.966); accuracy 0.843; sensitivity 0.829; specificity 0.854; positive predictive value 0.806; and negative predictive value 0.872.

Conclusion
In this prognostic study, we developed a machine-learning model that accurately predicted outcome after surgery for ASD. The model is built on routinely modifiable indicators, thereby facilitating its integration into clinical practice to promote optimized decision-making.

Cite this article: *Bone Joint J* 2025;107-B(3):337–345.

Correspondence should be sent to S. Lu; email: spinelu@163.com

doi:10.1302/0301-620X.107B3.
BJJ-2024-1220.R1 $2.00

Bone Joint J
2025;107-B(3):337–345.

Introduction
Adult spinal deformity (ASD) is highly prevalent among the elderly and has a significant impact on health-related quality of life (HRQoL).[1] When conservative management and physical therapy fail to provide relief, elective surgical intervention can be considered.[2] Although surgery can offer substantial benefits, postoperative recovery after operations for ASD remains arduous. Improvements in the Scoliosis Research Society-22r

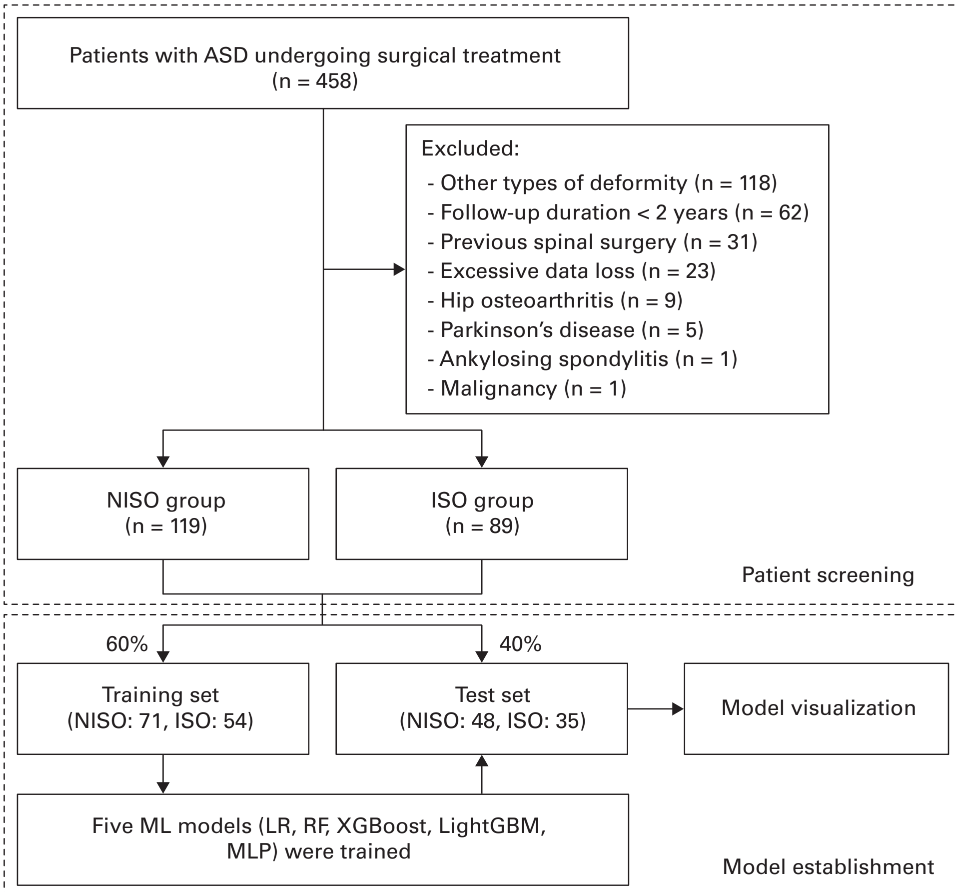


Fig. 1

Population selection and development of model. ASD, adult spinal deformity; ISO, ideal surgical outcome; LightGBM, light gradient boosting machine; LR, logistic regression; MLP, multilayer perceptron; NISO, no ideal surgical outcome; RF, random forest; XGBoost, extreme gradient boosting.

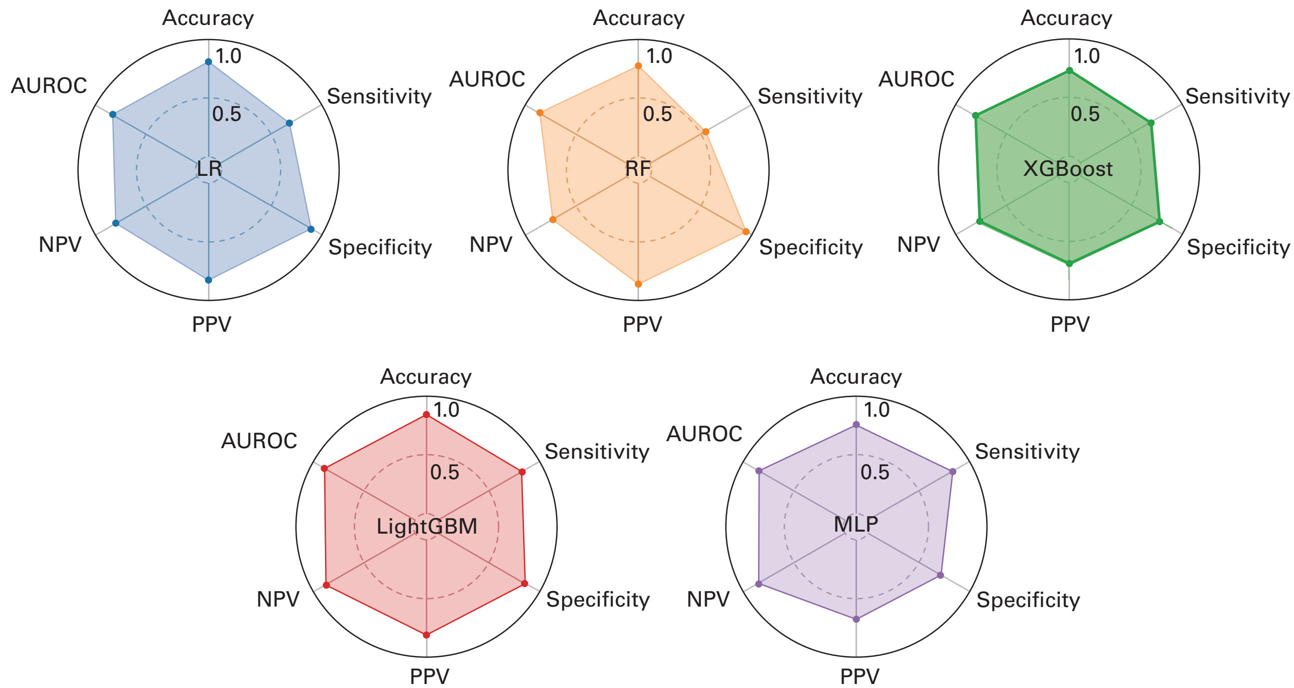
(SRS-22r)[3] score exceeding the minimal clinically important difference (MCID), a widely adopted indicator for assessing HRQoL in patients with ASD, are observed in only approximately half of those undergoing surgery.[4–6] Moreover, various complications, both medical and mechanical, may arise during the recovery period, affecting between 10% and 60% of patients.[7–10] Hence, achieving an ideal surgical outcome (ISO) – defined as an improvement in HRQoL reaching the MCID and the absence of any complications – remains a highly sought-after goal for surgeons.

The factors influencing surgical outcomes in patients with an ASD are multidimensional, encompassing psychological factors such as anxiety or depression,[11,12] physical factors such as frailty and osteoporosis,[13,14] as well as surgery-related factors such as realignment goals, fusion levels, and operative techniques.[15–17] Identifying the crucial factors that affect surgical outcome can help clinicians tailor treatment plans, provide more accurate patient counselling, and potentially improve postoperative outcome. Consequently, there exists a significant interest in predicting which patients will achieve ISO following surgery for ASD.

Although numerous studies have identified factors that contribute to improved prognosis,[11–17] the predictive power of traditional methods of statistical analysis used in these studies is limited, and major differences of opinion remain about the importance of selected features. Furthermore, only a small number of the identified features are modifiable, reducing the clinical applicability of previous findings. With recent developments, machine-learning (ML) techniques have emerged as a promising approach for predicting outcome.[18] ML offers considerable advantages in identifying complex, non-linear relationships without relying on traditional assumptions, and can highlight the most influential predictors through feature importance methods.[19] The purpose of this study was to develop and internally validate a ML model to explore the factors most closely related to the achievement of ISO in patients with an ASD.

Methods

Patient population. We retrospectively analysed 458 consecutive patients hospitalized with ASD at our centre between January 2016 and June 2022. The inclusion criteria were: age ≥



a

Model	Auroc (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
LR	0.830 (0.736 to 0.924)	0.807	0.686	0.896	0.828	0.796
RF	0.863 (0.781 to 0.945)	0.771	0.543	0.938	0.864	0.738
XGBoost	0.813 (0.717 to 0.908)	0.735	0.686	0.771	0.686	0.771
LightGBM	0.888 (0.810 to 0.966)	0.843	0.829	0.854	0.806	0.872
MLP	0.853 (0.765 to 0.941)	0.759	0.829	0.708	0.674	0.850

b

Fig. 2

Performance metrics of machine-learning models. a) Radar plots for the outcomes, b) Summary of performance metrics. AUROC, area under the receiver operating characteristic curve; LightGBM, light gradient boosting machine; LR, logistic regression; MLP, multilayer perceptron; PPV, positive predictive value; NPV, negative predictive value; RF, random forest; XGBoost, extreme gradient boosting.

18 years; radiological evidence of ASD, defined as at least one of the following: sagittal vertical axis (SVA) > 50 mm, pelvic tilt (PT) ≥ 25°, pelvic incidence minus lumbar lordosis (PI-LL) > 10°, or thoracic kyphosis (TK) > 60°; ≥ four-level posterior instrumented fusion from the sacrum; and a minimum follow-up period of two years. The exclusion criteria were as follows: previous spinal surgery; concomitant Parkinson's disease, ankylosing spondylitis, spinal tuberculosis, sepsis, or malignancy; combined with hip or knee osteoarthritis; and post-traumatic deformity, adult idiopathic scoliosis of the thoracic spine, or de novo lumbar scoliosis.[1] Of these, 208 patients (mean age 64.62 years (SD 8.21); 48 males (23.1%), 160 females (76.9%)) who had undergone surgery for ASD were eligible for inclusion in our primary analysis (Figure 1). A comparison of the sociodemographic, clinical, and radiological characteristics of the ISO and no ISO (NISO) groups is shown in Table I. The study was reported in line with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines.[20]

Outcomes. The primary outcome was the achievement of an ISO after surgery in patients with an ASD. ISO was defined by the fulfillment of two criteria: first, an improvement in HRQoL score exceeding the MCID, defined as an increase in SRS-22r > 0.94 from baseline to final follow-up;[21] and second, the absence of complications during follow-up, including medical complications such as wound infection, deep venous thrombosis, or delirium, and mechanical complications such as proximal junctional kyphosis (PJK),[22] proximal junctional failure (PJF),[23] or rod fractures.

Predictors. We used all potential predictors available in our clinical database (detailed descriptions of each variable are provided in the Supplementary Table i). In brief, 43 features (predictor variables) were initially included, covering patient sociodemographic characteristics, comorbidity, clinical characteristics, surgical details, and radiological characteristics. For preliminary feature selection, we applied three methods: least absolute shrinkage and selection operator (LASSO), recursive feature elimination with random forest (RF-RFE), and

Table I. Sociodemographic, clinical, and radiological characteristics of the study patients.

Characteristic	Total	NISO	ISO	p-value
n	208	119	89	-
Mean age, yrs (SD)	64.62 (8.21)	65.57 (8.01)	63.34 (8.35)	0.052*
Female sex, n (%)	160 (76.92)	88 (73.95)	72 (80.90)	0.239†
Mean height, cm (SD)	166.31 (10.50)	165.77 (10.61)	167.02 (10.37)	0.397*
Mean weight, kg (SD)	70.59 (16.59)	71.86 (16.65)	68.89 (16.46)	0.202*
Mean BMI, kg/m^2 (SD)	25.89 (7.02)	26.45 (6.74)	25.14 (7.36)	0.185‡
Mean symptom duration, mths (SD)	85.06 (107.04)	84.08 (104.41)	86.37 (111.06)	0.879*
Mean CCI (SD)	2.13 (1.49)	2.22 (1.48)	2.02 (1.51)	0.349‡
Osteoporosis, n (%)	38 (18.27)	33 (27.73)	5 (5.62)	< 0.001†
Frailty, n (%)	54 (25.96)	41 (34.45)	13 (14.61)	0.001†
Malnutrition, n (%)	73 (35.10)	53 (44.54)	20 (22.47)	< 0.001†
Depression, n (%)	21 (10.10)	18 (15.13)	3 (3.37)	0.005†
Anxiety, n (%)	19 (9.13)	11 (9.24)	8 (8.99)	0.950†
Currently smoker, n (%)	29 (13.94)	15 (12.61)	14 (15.73)	0.520†
Mean rTCSA (SD)	2.19 (0.56)	2.12 (0.55)	2.28 (0.56)	0.037‡
Mean rFCSA (SD)	1.15 (0.40)	1.03 (0.33)	1.31 (0.43)	< 0.001‡
Mean IVDD severity (SD)	6.61 (0.80)	6.66 (0.85)	6.54 (0.72)	0.306*
FPC, n (%)	61 (29.33)	47 (39.50)	14 (15.73)	< 0.001†
Surgical details				
Mean surgical levels, n (SD)	8.15 (2.72)	8.11 (2.58)	8.21 (2.91)	0.785*
UIV in the upper thoracic region, n (%)	37 (17.79)	17 (14.29)	20 (22.47)	0.127†
Injection of cement at UIV + 1, n (%)	81 (38.94)	41 (34.45)	40 (44.94)	0.125†
Mean operating time, mins (SD)	342.12 (86.13)	346.82 (78.25)	335.84 (95.76)	0.379*
Mean EBL, ml (SD)	902.84 (513.81)	878.74 (540.09)	935.06 (477.52)	0.435*
Mean intraoperative transfusion, ml (SD)	988.45 (645.23)	991.24 (669.84)	984.72 (614.50)	0.943*
Mean preoperative radiological measures (SD)				
TK, °	-23.08 (13.96)	-23.81 (14.60)	-22.10 (13.08)	0.384‡
TLK, °	-15.46 (15.78)	-14.93 (16.67)	-16.17 (14.56)	0.576*
LL, °	14.17 (14.81)	12.36 (15.48)	16.59 (13.57)	0.041‡
SS, °	17.30 (9.45)	17.34 (10.23)	17.24 (8.36)	0.938*
PT, °	31.45 (10.54)	30.76 (10.80)	32.37 (10.18)	0.277‡
PI, °	48.49 (9.79)	47.96 (9.87)	49.20 (9.69)	0.368‡
SVA, mm	109.37 (47.04)	113.70 (46.03)	103.59 (47.99)	0.125‡
TPA, °	31.05 (11.41)	31.70 (11.21)	30.18 (11.66)	0.344‡
Mean postoperative radiological measures (SD)				
TK, °	-29.34 (10.33)	-30.63 (10.07)	-27.62 (10.47)	0.037‡
TLK, °	-8.97 (9.74)	-9.34 (9.66)	-8.49 (9.89)	0.537‡
LL, °	33.00 (9.96)	31.50 (10.23)	35.01 (9.27)	0.012*
SS, °	26.25 (8.12)	24.50 (8.02)	28.59 (7.68)	< 0.001‡
Mean PT, °	23.48 (9.55)	25.13 (9.43)	21.27 (9.30)	0.004‡
PI, °	49.07 (9.61)	48.50 (9.93)	49.83 (9.16)	0.324‡
SVA, mm	46.80 (38.21)	51.13 (43.62)	41.00 (28.67)	0.045‡
TPA, °	19.25 (9.15)	21.07 (9.29)	16.83 (8.41)	< 0.001*
Realignment, n (%)				
PI-LL match	69 (33.17)	28 (23.53)	41 (46.07)	< 0.001†
PT match	98 (47.12)	40 (33.61)	58 (65.17)	< 0.001†
TPA match	106 (50.96)	49 (41.18)	57 (64.04)	< 0.001†
SAAS match	76 (36.54)	31 (26.05)	45 (50.56)	< 0.001†
Mean preoperative SRS-22r measures (SD)				
Pain	2.68 (0.62)	2.72 (0.60)	2.64 (0.64)	0.337‡
Function	2.80 (0.52)	2.79 (0.55)	2.82 (0.47)	0.632‡
Self-image	2.74 (0.53)	2.72 (0.50)	2.76 (0.58)	0.575*
Mental health	3.06 (0.51)	3.02 (0.52)	3.13 (0.48)	0.105*
Subtotal	2.82 (0.29)	2.81 (0.31)	2.84 (0.27)	0.505‡
Mean postoperative SRS-22r measures (SD)				
Pain	3.81 (0.46)	3.72 (0.43)	3.93 (0.47)	< 0.001‡
Function	3.81 (0.51)	3.74 (0.46)	3.91 (0.55)	0.022‡

Continued

Table I. Continued

Characteristic	Total	NISO	ISO	p-value
Self-image	4.03 (0.49)	3.87 (0.43)	4.24 (0.48)	< 0.001*
Mental health	3.88 (0.48)	3.80 (0.48)	3.98 (0.48)	0.006‡
Subtotal	3.88 (0.34)	3.78 (0.34)	4.01 (0.29)	< 0.001‡
Reached MCID, n (%)	129 (62.02)	40 (33.61)	89 (100.00)	< 0.001†
Postoperative adverse events, n (%)				
DVT	10 (4.81)	10 (8.40)	0 (0.00)	0.006§
Delirium	4 (1.92)	4 (3.36)	0 (0.00)	0.137§
Haematoma	10 (4.81)	10 (8.40)	0 (0.00)	0.006§
Pneumonia	6 (2.88)	6 (5.04)	0 (0.00)	0.039§
UTI	8 (3.85)	8 (6.72)	0 (0.00)	0.011§
Urinary retention	14 (6.73)	14 (11.76)	0 (0.00)	< 0.001†
Wound infection	18 (8.65)	18 (15.13)	0 (0.00)	< 0.001†
Pseudarthrosis	17 (8.17)	17 (14.29)	0 (0.00)	< 0.001†
PJK	47 (22.60)	47 (39.50)	0 (0.00)	< 0.001†
PJF	12 (5.77)	12 (10.08)	0 (0.00)	0.002†
Rod fractures	4 (1.92)	4 (3.36)	0 (0.00)	0.137§
Reoperation	20 (9.62)	20 (16.81)	0 (0.00)	< 0.001†
Mean follow-up, mths (SD)	29.56 (6.18)	28.84 (5.66)	30.52 (6.74)	0.053*

*Mann-Whitney U test.
†Chi-squared test.
‡Independent-samples *t*-test.
§Fisher's exact test.
CCI, Charlson Comorbidity Index; DVT, deep venous thrombosis; EBL, estimated blood loss; FPC, failure of pelvic compensation; IVDD, intervertebral disc degeneration; LL, lumbar lordosis; MCID, minimal clinically important difference; PI, pelvic incidence; PI-LL, pelvic incidence minus lumbar lordosis; PJF, proximal junctional failure; PJK, proximal junctional kyphosis; PT, pelvic tilt; rFCSA, relative functional cross-sectional area; rTCSA, relative total cross-sectional area; SAAS, sagittal age-adjusted score; SRS-22r, Scoliosis Research Society-22r; SS, sacral slope; SVA, sagittal vertical axis; TK, thoracic kyphosis; TLK, thoracolumbar kyphosis; TPA, T1 pelvic angle; UIV, upper instrumented vertebra; UTI, urinary tract infection.

the Boruta algorithm. The final set of input variables was determined by selecting the intersection of the results from these methods.

Model development. Five supervised ML algorithms were trained to develop the prediction model: logistic regression (LR), random forest (RF), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), and multilayer perceptron (MLP). The dataset was randomly split into training (60%) and test (40%) sets. Preprocessing steps involved centring and scaling of continuous variables and one-hot encoding (event/yes: 1, no event/no: 0) of categorical variables. For the variables of interest, missing data accounted for less than 5%. Consequently, a complete-case analysis approach was adopted, including only patients with no missing data for any of the variables.[24,25] Hyperparameters were tuned using a grid search with fivefold cross-validation (Supplementary Table ii). The optimal combination of hyperparameters was selected based on the highest area under the receiver operating characteristic curve (AUROC).

Model evaluation. The primary model evaluation metric was AUROC. Secondary performance metrics were accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Receiver operating characteristic (ROC) curves and precision-recall (PR) curves were plotted for data visualization. Clinical decision curve analysis (DCA) was undertaken to evaluate the clinical utility of the model.[26] To assess model robustness, we plotted calibration curves and calculated Brier scores.[27] Shapley additive explanations (SHAP) values were calculated for the prediction model to investigate feature importance.[28] All analyses were conducted using R v. 4.4.0 (R Foundation for Statistical Computing, Austria). The following R packages were used: *tidymodels*, *caret*, *mlbench*, *Boruta*, *glmnet*, and *fastshap*.

Statistical analysis. All analyses were performed using SPSS Statistics v. 29.0 (IBM, USA). The assumption of a normal distribution of the data was verified using the Shapiro-Wilk test. For continuous variables, comparison between groups was carried out using the independent-samples *t*-test or Mann-Whitney U test. The chi-squared test or Fisher's exact probability test was used to compare categorical variables between groups. Continuous variables were expressed as mean (SD). A p-value < 0.05 was considered to indicate statistical significance.

Results

Among the patients screened, a total of 89 (42.8%) achieved ISO after surgery.

Key variables. Three validated ML algorithms (LASSO, RF-RFE, Boruta) were applied to identify key variables from 43 features, yielding 17, 18, and 12 variables, respectively (Supplementary Table iii). Eight variables were selected by taking the intersection of the three results (Supplementary Figure a), including depression, osteoporosis, frailty, failure of pelvic compensation (FPC), relative functional cross-sectional area (rFCSA) of paraspinal muscles, postoperative sacral slope (SS), postoperative PT match, and postoperative match per sagittal age-adjusted score (SAAS).

Model performance. Among the five ML models assessed using the test set data, LightGBM had the best performance in

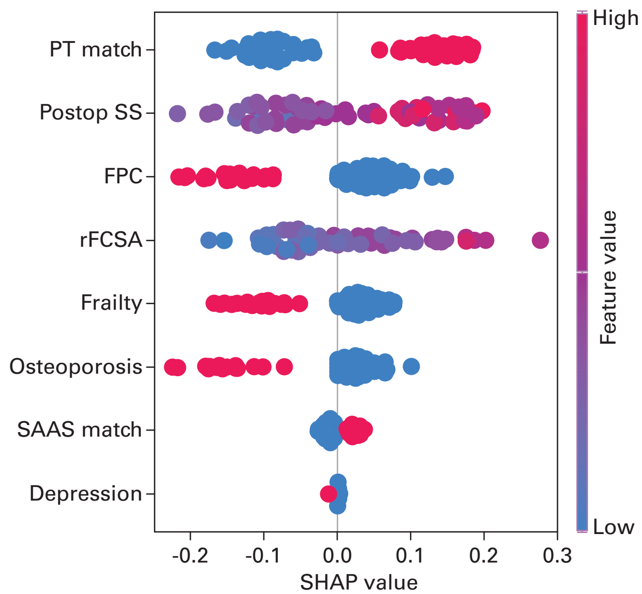


Fig. 3

Shapley additive explanations (SHAP) summary plot of variable importance for the ideal surgical outcome prediction model. FPC, failure of pelvic compensation; PT, pelvic tilt; rFCSA, relative total cross-sectional area; SAAS, sagittal age-adjusted score; SS, sacral slope.

predicting ISO achievement in ASD patients (AUROC 0.888 (95% CI 0.810 to 0.966)). By comparison, the other models had the following AUROCs: LR 0.830 (95% CI 0.736 to 0.924); RF 0.863 (95% CI 0.781 to 0.945); XGBoost 0.813 (95% CI 0.717 to 0.908); and MLP 0.853 (95% CI 0.765 to 0.941). The net benefit of the LightGBM model surpassed that of other models across a wide range of threshold probabilities in the DCA (Supplementary Figure b), indicating its superior clinical utility. Model performance results are summarized in Figure 2. The LightGBM had an accuracy of 0.843; sensitivity of 0.829; specificity of 0.854; PPV of 0.806; and NPV of 0.872. Calibration plots in Supplementary Figure c illustrate good agreement between predicted and observed event probabilities for LightGBM, with a Brier score of 0.136.

Feature importance. By SHAP values, the top three important features for achieving an ideal surgical outcome were a match in PT following surgery, higher postoperative SS, and the absence of FPC (Figure 3). Additionally, the absence of osteoporosis, frailty, and depression, as well as higher relative functional cross-sectional area and sagittal age-adjusted score match, were also beneficial for achieving ISO. Supplementary Figures d to f show representative SHAP force plots illustrating these findings.

Discussion

In this prognostic study, we developed and validated a ML model to predict the achievement or non-achievement of an ideal treatment outcome in patients with ASD after surgery. The model incorporated eight key features: depression, frailty, osteoporosis, rFCSA, FPC, postoperative SS, PT match, and SAAS match. Additionally, the model showed strong discriminative performance, with an AUROC of 0.89. Our model has the potential to support clinical decision-making throughout a patient's perioperative course by facilitating the assessment of risks to the individual.

Overall, the associations identified between significant predictive variables and surgical outcomes align with previous literature. For instance, concomitant frailty, depression, and osteoporosis have all been widely associated with higher rates of adverse events in patients with ASD who undergo spinal fusion surgery.[29–31] Paraspinal muscle atrophy has been proven to increase the risk of postoperative mechanical complications, including PJK and PJF.[32,33] FPC, characterized by substantial anterior deviation of SVA with minimal pelvic retroversion compensation, is a well-established risk factor for suboptimal recovery following ASD surgery, both clinically and radiologically.[34,35] Moreover, restoring spinal sagittal alignment improves surgical outcomes in patients with ASD, as they lose flexibility in the fused spinal segment and are more prone to symptoms related to sagittal malalignment because of a reduction in their ability to maintain a balanced standing posture.[36–38] Consequently, all the key features enrolled in this study are closely related to surgical prognosis for ASD, which indirectly proves the rationality of our model and the reliability of the predicted results.

Several previous studies have derived predictive models for HRQoL or postoperative complications in ASD patients. Ames et al[39] developed a MCID prediction model and proposed that patients with worse preoperative baseline HRQoL (including SRS-22r, Oswestry Disability Index, and 36-Item Short-Form Health Survey questionnaire) achieved the greatest improvements after surgery. Similarly, Ryu et al[40] reported that postoperative global sagittal balance, as measured by SVA, played a vital role in predicting unplanned reoperation after corrective surgery for ASD. Additionally, Lee et al[41] developed a PJK prediction model, identifying age, BMI, deformity type based on SRS-Schwab criteria, baseline PI, and postoperative proximal junctional angle as independent predictors. However, few features reported in previous models are modifiable. In contrast, by leveraging three ML algorithms, we identified eight key predictors for ISO in our model, involving preoperative comorbidities, paraspinal muscle mass, pelvic compensation capacity, and postoperative spinal sagittal morphology. The majority of the selected indicators in our prediction model can be optimized by targeted intervention and tailored surgical planning preoperatively, enhancing the generalizability of our model.

A critical implication of our prediction model is to serve as a reference for perioperative management in patients treated surgically for ASD. Based on our findings, screening and addressing baseline comorbidities should be the focus of preoperative optimization of a patient's condition. Tools like dual-energy x-ray absorptiometry and CT-Hounsfield units are effective for evaluating bone quality.[42] For ASD patients with osteoporosis, initiating anabolic therapy, such as teriparatide, one to three months before surgery can contribute to enhancing spinal fusion and mitigating the incidence of mechanical complications.[43,44] With regard to depression, there is no universal consensus on the most effective screening tool. The Zung's Self-Rating Depression Scale might be a preferable selection considering

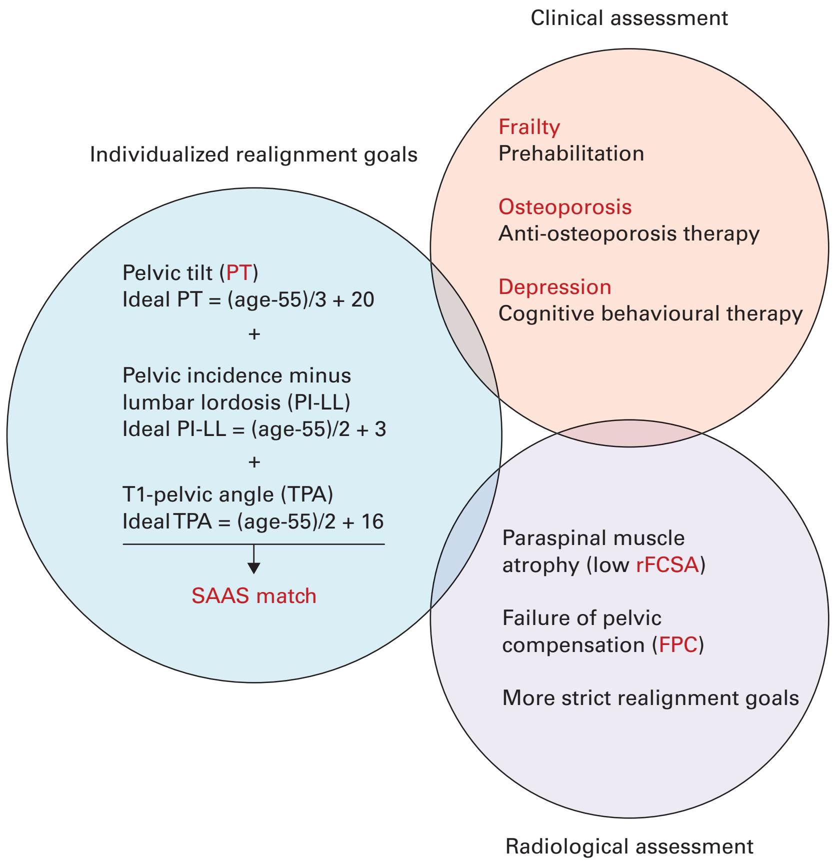


Fig. 4

Key factors for preoperative optimization of patients with adult spinal deformity. The sagittal age-adjusted score (SAAS) system is cited from Lafage et al.[15] rFCSA, relative functional cross-sectional area.

its wide clinical application.[45,46] Patients with depression should be referred to the behavioural health services and monitored for improvement in their symptoms before proceeding with ASD surgery.[47] As for frailty, the Fried frailty phenotype is an easy and reliable screening tool.[48] Although there is a lack of sufficient evidence, prehabilitation could be beneficial for the improvement of functional status preoperatively and for priming patients with frailty to withstand the stressors associated with deformity correction.[47]

Moreover, our findings indicate that thorough preoperative surgical planning to ensure restoration of sagittal alignment is another key element that requires attention. In this study, sagittal realignment was assessed using the SAAS system, which is constructed based on three parameters routinely used for sagittal alignment evaluation (PT, PI-LL, TPA) and takes into account the natural degeneration of the spine with age.[49] However, the optimal corrective goal for ASD has still not been established. For instance, a proportioned spinopelvic state based on the global alignment and proportion (GAP) score was shown to be related to lower complication rates in ASD patients by Yilgor et al.[50] By contrast, the Roussouly algorithm was considered by Gessara et al[51] to be a valuable reference for ASD surgery to reduce postoperative mechanical complications. A recent study by Park et al[52] examined the association between four criteria (SAAS, GAP, SRS-Schwab modifier, and the Roussouly algorithm) and surgical outcomes, finding that correction according to the SAAS and restoring the Roussouly type most significantly improved prognosis. Despite these differing strategies, customizing surgical strategies to individual patients remains a core principle to which surgeons should adhere. Based on our findings, setting personalized realignment goals based on SAAS parameters may enhance surgical outcomes for ASD patients; however, further research is still warranted to address the remaining discrepancies among realignment strategies.

Our study has several limitations. First, it is a retrospective study with a limited sample size. Nonetheless, we applied strict inclusion and exclusion criteria and thoroughly preprocessed all data to enhance consistency and reduce potential biases. Second, as this study was conducted at a single centre, the external validity of our findings is restricted, potentially affecting the generalizability and robustness of the model. To validate these results and further refine the prediction model, a prospective, multicentre study with a larger and more diverse patient population will be necessary. Additionally, future studies could benefit from integrating multimodal data sources, such as imaging, genetic, and clinical data, to enhance model performance and provide a more comprehensive understanding of the factors influencing outcomes. This approach may improve the model's predictive accuracy and robustness across diverse patient populations.

In summary, we developed and validated a ML model with high discriminative performance for predicting the achievement of ideal outcomes after corrective surgery for ASD. The model is built on routinely modifiable indicators, enhancing its potential for clinical application. Our finding highlights the importance of implementing multidisciplinary and individualized management strategies for patients with ASD (Figure 4); this should be a focal point in clinical practice and future studies. Prospective clinical validation of our prediction model is warranted to evaluate its practical utility and feasibility for integration into clinical workflows.

Take home message

- This study demonstrates the use of machine-learning algorithms, particularly light graident-boosting machine, to accurately predict the likelihood of achieving an ideal surgical outcome in patients undergoing adult spinal deformity surgery (ASD). This approach could enhance preoperative decision-making and patient counselling.
- Several modifiable factors, including depression, osteoporosis, frailty, pelvic compensation capacity, paraspinal muscle degeneration, and sagittal alignment restoration, were identified as significant predictors of surgical success.
- These findings highlight the importance of a multidisciplinary assessment and individualized interventions to optimize outcomes in patients with ASD.

Supplementary material

Detailed description of the variables involved in this study, the results of variable selection, as well as the area under the curve, precision-recall curve, decision curve analysis plot, and calibration curve, along with Shapley additive explanations (SHAP) force plots for patients with different SHAP values.

References

1. **Diebo BG**, **Shah NV**, **Boachie-Adjei O**, **et al**. Adult spinal deformity. *Lancet*. 2019;394(10193):160–172.
2. **Ames CP**, **Scheer JK**, **Lafage V**, **et al**. Adult spinal deformity: epidemiology, health impact, evaluation, and management. *Eval and Manag Spine Deform*. 2016;4(4):310–322.
3. **Asher M**, **Min Lai S**, **Burton D**, **Manna B**. The reliability and concurrent validity of the Scoliosis Research Society-22 patient questionnaire for idiopathic scoliosis. *Spine (Phila Pa 1976)*. 2003;28(1):63–69.
4. **Arima H**, **Hasegawa T**, **Yamato Y**, **et al**. Factors associated with improved quality of life outcomes in patients undergoing surgery for adult spinal deformity. *Spine (Phila Pa 1976)*. 2021;46(6):E384–E391.
5. **Wang D**, **Wang W**, **Wang Y**, **et al**. Identification and impact of failure of pelvic compensation in patients with adult spinal deformity. *Spine J*. 2024;24(11):2124–2134.
6. **Mohanty S**, **Lai C**, **Mikhail C**, **et al**. Outcomes and reoperation rates of adult spinal deformity patients with baseline sagittal malalignment based on the cranial axis to the hip at 2 years postoperatively. *J Neurosurg Spine*. 2023;39(3):301–310.
7. **Alas H**, **Passias PG**, **Brown AE**, **et al**. Predictors of serious, preventable, and costly medical complications in a population of adult spinal deformity patients. *Spine J*. 2021;21(9):1559–1566.
8. **Bortz C**, **Pierce KE**, **Brown A**, **et al**. Frequency and implications of concurrent complications following adult spinal deformity corrective surgery. *Spine (Phila Pa 1976)*. 2021;46(21):E1155–E1160.
9. **Alvarez Reyes A**, **Jack AS**, **Hurlbert RJ**, **Ramey WL**. Complications in the elderly population undergoing spinal deformity surgery: a systematic review and meta-analysis. *Global Spine J*. 2022;12(8):1934–1942.
10. **Burke JF**, **Scheer JK**, **Lau D**, **et al**. Failure in adult spinal deformity surgery: a comprehensive review of current rates, mechanisms, and prevention strategies. *Spine (Phila Pa 1976)*. 1976;47(19):1337–1350.
11. **Agarwal N**, **Letchuman V**, **Lavadi RS**, **et al**. What is the effect of preoperative depression on outcomes after minimally invasive surgery for adult spinal deformity? A prospective cohort analysis. *J Neurosurg Spine*. 2024;40(5):602–610.
12. **Shah I**, **Wang C**, **Jain N**, **Formanek B**, **Buser Z**, **Wang JC**. Postoperative complications in adult spinal deformity patients with a mental illness undergoing reconstructive thoracic or thoracolumbar spine surgery. *Spine J*. 2019;19(4):662–669.
13. **Williamson TK**, **Passfall L**, **Ihejirika-Lomedico R**, **et al**. Assessing the influence of modifiable patient-related factors on complication rates after adult spinal deformity surgery. *Bone Joint J*. 2022;104-B(11):1249–1255.
14. **Pierce KE**, **Passias PG**, **Alas H**, **et al**. Does patient frailty status influence recovery following spinal fusion for adult spinal deformity? An analysis of patients with 3-year follow-up *Spine (Phila Pa 1976)*. 2020;45(7):E397–E405.
15. **Lafage R**, **Smith JS**, **Elysee J**, **et al**. Sagittal age-adjusted score (SAAS) for adult spinal deformity (ASD) more effectively predicts surgical outcomes and proximal junctional kyphosis than previous classifications. *Spine Deform*. 2022;10(1):121–131.
16. **Park SJ**, **Lee CS**, **Park JS**, **Lee KJ**. Should thoracolumbar junction be always avoided as upper instrumented vertebra in long instrumented fusion for adult spinal deformity? Risk factor analysis for proximal junctional failure. *Spine (Phila Pa 1976)*. 2020;45(10):686–693.
17. **Passias PG**, **Williamson TK**, **Joujon-Roche R**, **et al**. The impact of lumbopelvic realignment versus prevention strategies at the upper-instrumented vertebra on the rates of junctional failure following adult spinal deformity surgery. *Spine (Phila Pa 1976)*. 2024;49(6):E72–E78.
18. **Shamout F**, **Zhu T**, **Clifton DA**. Machine learning for clinical outcome prediction. *IEEE Rev Biomed Eng*. 2021;14:116–126.
19. **Baştanlar Y**, **Ozuysal M**. Introduction to machine learning. *Methods Mol Biol*. 2014;1107:105–128.
20. **Collins GS**, **Reitsma JB**, **Altman DG**, **Moons KGM**. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
21. **Yuksel S**, **Ayhan S**, **Nabiyev V**, **et al**. Minimum clinically important difference of the health-related quality of life scales in adult spinal deformity calculated by latent class analysis: is it appropriate to use the same values for surgical and nonsurgical patients? *Spine J*. 2019;19(1):71–78.
22. **Glattes RC**, **Bridwell KH**, **Lenke LG**, **Kim YJ**, **Rinella A**, **Edwards C**. Proximal junctional kyphosis in adult spinal deformity following long instrumented posterior spinal fusion: incidence, outcomes, and risk factor analysis. *Spine (Phila Pa 1976*). 2005;30(14):1643–1649.
23. **Lafage R**, **Schwab F**, **Glassman S**, **et al**. Age-adjusted alignment goals have the potential to reduce PJK. *Spine (Phila Pa 1976)*. 2017;42(17):1275–1282.
24. **Hughes RA**, **Heron J**, **Sterne JAC**, **Tilling K**. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int J Epidemiol*. 2019;48(4):1294–1304.
25. **Ross RK**, **Breskin A**, **Westreich D**. When is a complete-case approach to missing data valid? The importance of effect-measure modification. *Am J Epidemiol*. 2020;189(12):1583–1589.
26. **Vickers AJ**, **Elkin EB**. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565–574.
27. **Redelmeier DA**, **Bloch DA**, **Hickam DH**. Assessing predictive accuracy: how to compare Brier scores. *J Clin Epidemiol*. 1991;44(11):1141–1146.
28. **Lundberg SM**, **Lee S-I**. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30.
29. **Diebo BG**, **Lavian JD**, **Murray DP**, **et al**. The impact of comorbid mental health disorders on complications following adult spinal deformity surgery with minimum 2-year surveillance. *Spine (Phila Pa 1976)*. 2018;43(17):1176–1183.
30. **Wang S-K**, **Wang Q-J**, **Wang P**, **et al**. The impact of frailty on clinical outcomes of older patients undergoing enhanced recovery after lumbar fusion surgery: a prospective cohort study. *Int J Surg*. 2024;110(8):4785–4795.
31. **Gupta A**, **Cha T**, **Schwab J**, **et al**. Osteoporosis increases the likelihood of revision surgery following a long spinal fusion for adult spinal deformity. *Spine J*. 2021;21(1):134–140.
32. **Park JS**, **Cho KJ**, **Kim JS**, **Park SJ**, **Baek H**. Sarcopenia in paraspinal muscle as a risk factor of proximal junctional kyphosis and proximal junctional failure after adult spinal deformity surgery. *J Neurosurg Spine*. 2024;40(3):324–330.
33. **Pennington Z**, **Cottrill E**, **Ahmed AK**, **et al**. Paraspinal muscle size as an independent risk factor for proximal junctional kyphosis in patients undergoing thoracolumbar fusion. *J Neurosurg Spine*. 2019;31(3):380–388.

34. **Kwon O**, **Lee S**, **Park SM**, **Yeom JS**, **Kim HJ**. A complement type to SRS-Schwab adult spinal deformity classification: the failure of pelvic compensation. *Spine (Phila Pa 1976)*. 2022;47(18):1295–1302.
35. **Ferrero E**, **Vira S**, **Ames CP**, **et al**. Analysis of an unexplored group of sagittal deformity patients: low pelvic tilt despite positive sagittal malalignment. *Eur Spine J*. 2016;25(11):3568–3576.
36. **Takemoto M**, **Boissière L**, **Vital J-M**, **et al**. Are sagittal spinopelvic radiographic parameters significantly associated with quality of life of adult spinal deformity patients? Multivariate linear regression analyses for pre-operative and short-term post-operative health-related quality of life. *Eur Spine J*. 2017;26(8): 2176–2186.
37. **Kyrölä K**, **Repo J**, **Mecklin J-P**, **Ylinen J**, **Kautiainen H**, **Häkkinen A**. Spinopelvic changes based on the simplified SRS-Schwab adult spinal deformity classification: relationships with disability and health-related quality of life in adult patients with prolonged degenerative spinal disorders. *Spine (Phila Pa 1976*). 2018;43(7): 497–502.
38. **Passias PG**, **Passfall L**, **Williamson TK**, **et al**. The clinical impact of failing to achieve ideal proportional realignment in adult spinal deformity patients. *Spine (Phila Pa 1976)*. 2022;47(14):995–1002.
39. **Ames CP**, **Smith JS**, **Pellisé F**, **et al**. Development of deployable predictive models for minimal clinically important difference achievement across the commonly used health-related quality of life instruments in adult spinal deformity surgery. *Spine (Phila Pa 1976)*. 2019;44(16):1144–1153.
40. **Ryu S-J**, **So J-Y**, **Ha Y**, **et al**. Risk factors for unplanned reoperation after corrective surgery for adult spinal deformity. *Bone Joint Res*. 2023;12(4):245–255.
41. **Lee C-H**, **Jo D-J**, **Oh JK**, **et al**. Development and validation of an online calculator to predict proximal junctional kyphosis after adult spinal deformity surgery using machine learning. *Neurospine*. 2023;20(4):1272–1280.
42. **Sardar ZM**, **Coury JR**, **Cerpa M**, **et al**. Best practice guidelines for assessment and management of osteoporosis in adult patients undergoing elective spinal reconstruction. *Spine (Phila Pa 1976)*. 1976;47(2):128–135.
43. **Ohtori S**, **Inoue G**, **Orita S**, **et al**. Comparison of teriparatide and bisphosphonate treatment to reduce pedicle screw loosening after lumbar spinal fusion surgery in postmenopausal women with osteoporosis from a bone quality perspective. *Spine (Phila Pa 1976)*. 2013;38(8):E487–E492.
44. **Ebata S**, **Takahashi J**, **Hasegawa T**, **et al**. Role of weekly teriparatide administration in osseous union enhancement within six months after posterior or transforaminal lumbar interbody fusion for osteoporosis-associated lumbar degenerative disorders: a multicenter, prospective randomized study. *J Bone Joint Surg Am*. 2017;99-A(5):365–372.
45. **Cui P**, **Liu T**, **Wang Z**, **et al**. Comprehensive geriatric assessment to predict adverse events in elderly patients undergoing posterior lumbar fusion surgery. *Eur Spine J*. 2024;33(1):274–281.
46. **Wang SK**, **Cui P**, **Wang DF**, **Wang P**, **Kong C**, **Lu SB**. Preoperative Zung depression scale predicts outcomes in older patients undergoing short-segment fusion surgery for degenerative lumbar spinal disease. *Eur Spine J*. 2023;32(2):718–726.
47. **Katiyar P**, **Reyes J**, **Coury J**, **Lombardi J**, **Sardar Z**. Preoperative optimization for adult spinal deformity surgery: a systematic review. *Spine (Phila Pa 1976)*. 1976;49(5):304–312.
48. **Fried LP**, **Tangen CM**, **Walston J**, **et al**. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146–56.
49. **Lafage R**, **Schwab F**, **Challier V**, **et al**. Defining spino-pelvic alignment thresholds: should operative goals in adult spinal deformity surgery account for age? *Spine (Phila Pa 1976)*. 1976;41(1):62–68.
50. **Yilgor C**, **Sogunmez N**, **Boissiere L**, **et al**. Global Alignment and Proportion (GAP) score: development and validation of a new method of analyzing spinopelvic alignment to predict mechanical complications after adult spinal deformity surgery. *J Bone Joint Surg Am*. 2017;99-A(19):1661–1672.
51. **Gessara A**, **Patel MS**, **Estefan M**, **et al**. Restoration of the sagittal profile according to the Roussouly classification reduces mechanical complications and revision surgery in older patients undergoing surgery for adult spinal deformity (ASD). *Eur Spine J*. 2024;33(2):563–570.
52. **Park SJ**, **Lee CS**, **Park JS**, **Jeon CY**, **Ma CH**. A validation study of four preoperative surgical planning tools for adult spinal deformity surgery in proximal junctional kyphosis and clinical outcomes. *Neurosurgery*. 2023;93(3):706–716.

Author information:
D. Wang, MD, Orthopaedic Surgeon
Q. Wang, MD, Orthopaedic Surgeon
P. Cui, MD, Orthopaedic Surgeon
S. Wang, MD, Orthopaedic Surgeon
D. Han, MD, Orthopaedic Surgeon
X. Chen, MD, PhD, Orthopaedic Surgeon
S. Lu, MD, Orthopaedic Surgeon
Department of Orthopedics & Elderly Spinal Surgery, Xuanwu Hospital of Capital Medical University, National Clinical Research Center for Geriatric Diseases, Beijing, China.

Author contributions:
D. Wang: Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing.
Q. Wang: Data curation, Formal analysis, Methodology, Software, Validation, Visualization.
P. Cui: Data curation, Methodology, Resources, Software.
S. Wang: Visualization, Methodology, Software.
D. Han: Methodology, Software.
X. Chen: Conceptualization, Supervision, Writing – review & editing.
S. Lu: Conceptualization, Supervision, Writing – review & editing.

X. Chen and S. Lu contributed equally to this work.

D. Wang and Q. Wang are joint first authors.

Funding statement:
The authors disclose receipt of the following financial or material support for the research, authorship, and/or publication of this article: this study was supported by post-subsidy funds from the National Clinical Research Center, Ministry of Science and Technology of China (No. 303-01-001-0272-05), the National Key Research and Development Program of China (No.2020YFC2004900), R&D Program of Beijing Municipal Education Commission (22JG0059), the Key Science and Technology Project of Beijing Municipal Education Commission (KZ20231002537), and the Beijing Natural Science Foundation (7242065).

ICMJE COI statement:
S. Lu reports post-subsidy funds from the National Clinical Research Center, Ministry of Science and Technology of China (No. 303-01-001-0272-05) to cover the open access fee for this article.

Data sharing:
The data that support the findings for this study are available to other researchers from the corresponding author upon reasonable request.

Acknowledgements:
All authors listed meet the authorship criteria according to the latest guidelines of the International Committee of Medical Journal Editors. All authors are in agreement with the manuscript.

Ethical review statement:
All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consents were obtained from all participants in the study. All experiments were performed in accordance with relevant guidelines and regulations. All experimental protocols were approved by the ethics committee of the Xuanwu Hospital Capital Medical University.

Open access funding:
The open access fee for this article was provided by the National Clinical Research Center, Ministry of Science and Technology of China (No. 303-01-001-0272-05).

This article was primary edited by A. C. Ross.